

## ANNOTATIONSGUIDELINES ZU PART OF SPEECH (POS)

### A. Grundsätzliches:

Auch das PoS-Tagging findet mithilfe des Tools CAB (Cascaded Analysis Broker), das vom DTA (Deutsches Text Archiv) entwickelt wurde, statt. Dieses automatische PoS-Tagging wird manuell kontrolliert und korrigiert. Dabei muss die Wortart des Tokens auf der PoS-Ebene mit der des Lemmas übereinstimmen. D.h. die Korrektur des Layers *PoS* kann erst nach der Korrektur der Layer *Norm* und *Lemma* vorgenommen werden.

Das PoS-Tagging ist NICHT als freie Annotation konzipiert, sondern es wurde ein geschlossenes Tagset (STTS) unter dem Layer POS verwendet.

Literatur:

Rehbein, J. (2014): Wortartenannotation im Kiezdeutschkorpus (KiDKo 1.0) – Draft. In: Kiezdeutschkorpus – KiDKo: Haupt- und Ergänzungskorpus. <http://www.kiezdeutschkorpus.de/de/kidko-haupt-und-ergaenzungskorpus> [zuletzt abgerufen am 11.06.2018].

Schiller, A./Teufel, S./Stöckert, C./Thielen, C. (1999): Guidelines für das Tagging deutscher Textcorpora mit STTS. (Kleines und großes Tagset). Universität Stuttgart, Institut für maschinelle Sprachverarbeitung; Universität Tübingen, Seminar für Sprachwissenschaft.

Westpfahl, S. et al. (2017): STTS 2.0 Guidelines für die Annotation von POS-Tags für Transkripte gesprochener Sprache in Anlehnung an das Stuttgart Tübingen Tagset (STTS). Mannheim.

### B. Tagset und Beispiele:

1.	Das Tagset findet sich am Ende dieser Datei.
2.	In einigen Fällen muss, um die syntaktische Funktion eines Token zuordnen zu können, die gesamte syntaktische Einheit betrachtet werden. Diese ist häufig durch die den Vers gemachte Segmentierung nicht innerhalb einer Zeile zu finden, sodass hier mehrere Zeilen betrachtet werden müssen.
3.	Umgang mit fremdsprachlichen Token/Phrasen/Sätzen:  a) Fremdsprachliche Passagen werden auf der Norm-Ebene nicht angepasst, d.h. sie werden nicht übersetzt. Die Lemmatisierung findet in der Form statt, dass das ursprüngliche Token und das Lemma übereinstimmen.  b) Das PoS-Tagging findet wie folgt statt: I. Lateinische Token/Phrasen/Sätze werden mit FM.la getaggt.

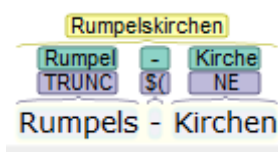
	<p>Ausnahme: Horribilicribrifax. Hier werden auch lateinische Passagen mit FM.xy getaggt (s. III.).</p> <p>II. Hebräische Token/Phrasen/Sätze werden mit FM.he getaggt.</p> <p>III. Alle anderen Sprachen (vornehmlich Französisch) werden mit FM.xy getaggt, da nicht sichergestellt werden kann, ob es sich um ‚echtes‘ Französisch handelt oder (z.B. durch Verballhornungen wie im Horribilicribrifax) eine vom damaligen Sprachstandard abweichende und nicht in der Sprachgemeinschaft verbreitete Variante genutzt wird.</p> <p>IV. Ausnahme: Treten fremdsprachliche Token singularär auf, d.h. sie sind in deutsche Passagen eingebettet, gilt Folgendes: Enthalten die fremdsprachlichen Token deutsch- sprachige Flexionsendungen, dann wird das fremdsprachliche Token in der entsprechenden Wortart getaggt. Sind keine deutschsprachigen Flexionsendungen zu erkennen, wird das Tag FM.xy verwendet (s. Einzelbeispiele).</p>
--	---

## Beispiele

**1) Komposita**, die durch einen Bindestrich getrennt werden oder auseinander geschrieben sind und deren Bestandteile je zwei Token darstellen, werden auf der PoS-Ebene wie folgt annotiert: Dem Erstglied wird immer das Tag TRUNC zugewiesen und dem Zweitglied die jeweilige vorherrschende Wortart.

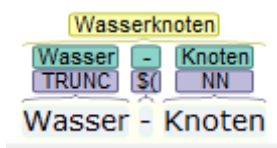
### Nominalkomposita

Beispiel 1: Sqentz



Rumpel → TRUNC und Kirche → NE

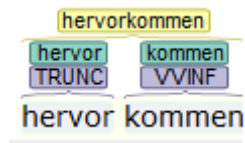
Beispiel 2: Sqentz



Wasser → TRUNC und Knoten → NN

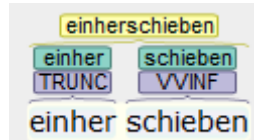
## Verbalkomposita

Beispiel 3: Squentz



hervor → TRUNC und kommen → VVINF

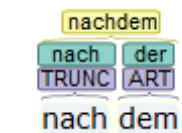
Beispiel 4: Squentz



einher → TRUNC und schieben → VVINF

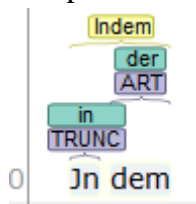
## Komposita, die eine Konjunktion ergeben

Beispiel 5: Majuma



nach → TRUNC und der → ART

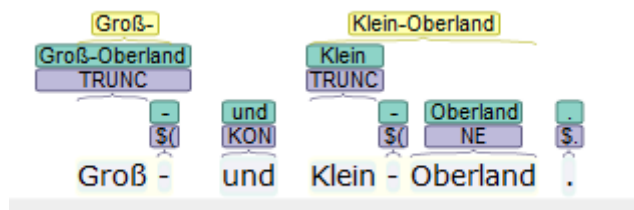
Beispiel 6: Catharina

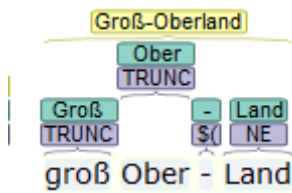


in → TRUNC und der → ART

Weist das Kompositum mehr als zwei Bestandteile auf, wird wie folgt annotiert: Das "erste" Erstglied erhält das Tag TRUNC, das "zweite" Erstglied ebenfalls und das Zweiglied wird mit der jeweiligen vorherrschenden Wortart getaggt.

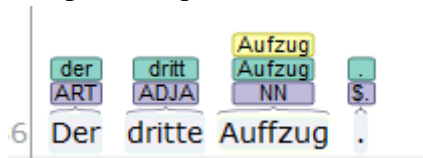
Beispiel 7: Squentz





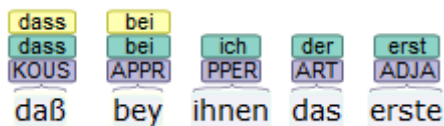
2) **Ordinalzahlen** (*erster, dritte, zehnter*) werden wie Adjektive behandelt und auch dementsprechend getaggt.

Beispiel 8: Squentz



dritt → ADJA

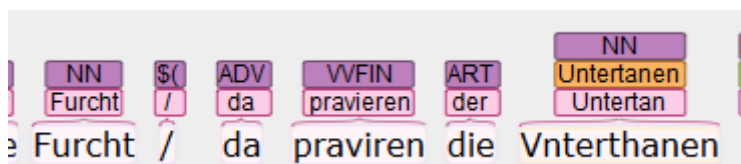
Beispiel 9: Squentz



erst → ADJA

3) **Fremdsprachliche Token, die in deutschsprachigen Text eingebettet sind** und eine deutschsprachige Flexionsendung aufweisen, werden mit der entsprechenden Wortart getaggt.

Beispiel 10: Verliebttes Gespenste



präviren → VVFIN

### C. Technisches Vorgehen:

1.	Einloggen in INCEption mit Benutzerkennung.
2.	Gryphius- bzw. Dramen-Projekt auswählen und auf <i>Annotation</i> klicken.
3.	Das Dokument öffnen, für dessen Korrektur/Annotation man eingeteilt wurde (Dokument anklicken).

4.	Gegebenenfalls unter <i>Settings</i> (Zahnrad-Symbol) für den Layer <i>Part of Speech</i> ein Häkchen setzen. In der Regel werden alle Layer automatisch ausgewählt.
5.	An der rechten Seite den Layer <i>Part of Speech</i> auswählen.
6.	Da alle Token mit einem <i>PoS</i> -Tag versehen wurden, kann immer die schon vorhandene Annotation über dem zu korrigierenden Token auf dem Layer <i>Part of Speech</i> per Doppelklick ausgewählt werden, um diese zu korrigieren. Der Eintrag wird orangefarben markiert, sodass die Auswahl ersichtlich wird.
7.	Auf der rechten Seite erscheint das Annotationsfeld <i>XPOS</i> . Darin wird das korrekte Tag aus dem Dropdown-Menü ausgewählt, sodass es auch über dem jeweiligen Token erscheint.
9.	Gegebenenfalls auf <i>Clear</i> klicken, um den Layer zu schließen, oder das nächste zu korrigierende Token auf dem Layer <i>Part of Speech</i> per Doppelklick auswählen.

Die Guidelines entstammen: Eggert, Lisa; Müller, Melissa (2021): „POS – Part of Speech“. In: „Guidelines. Interaktionale Sprache bei Gryphius“

URL: <https://gryphiusprojekt.wordpress.com/guidelines/pos-part-of-speech-2/>

**Tagset – STTS / STTS 2.0 (mit Ergänzungen aus dem Gryphius-Projekt)<sup>1</sup>**

Tag	Beschreibung	Beispiel
ADJA	attributives Adjektiv	[das] große [Haus]
ADJD	adverbiales oder prädikatives Adjektiv	[er fährt] schnell, [er ist] schnell
ADV	Adverb	schon, bald, doch
APPR	Präposition, Zirkumposition links	in [der Stadt], von [jetzt an]
APPRART	Präposition mit Artikel	im [Haus], zur [Sache]
APPO	Postposition	[ihm] zufolge, [der Sache] wegen
APZR	Zirkumposition rechts	[von jetzt] an
ART	Bestimmter oder unbestimmter Artikel	der, die, das, ein, eine
CARD	Kardinalzahl	zwei [Männer], [im Jahre] 1994
FM	Fremdsprachliches Material	[Bildet euch doch nicht dergleichen Gedancken ein /] absit injuria!
ITJ	Interjektion	mhm, ach, tja
KOUI	unterordnende Konjunktion mit ``zu" und Infinitiv	um [zu leben], anstatt [zu fragen]
KOUS	unterordnende Konjunktion mit Satz	weil, dass, damit, wenn, ob
KON	nebenordnende Konjunktion	und, oder, aber
KOKOM	Vergleichskonjunktion	als, wie
NN	Normales Nomen	Tisch, Herr, [das] Reisen
NE	Eigennamen	Hans, Hamburg, HSV
PDS	substituierendes Demonstrativpronomen	dieser, jener
PDAT	attribuierendes Demonstrativpronomen	jener [Mensch]
PIS	substituierendes Indefinitpronomen	keiner, viele, man, niemand
PIAT	attribuierendes Indefinitpronomen ohne Determiner	kein [Mensch], irgendein [Glas]
PIDAT	attribuierendes Indefinitpronomen mit Determiner	[ein] wenig [Wasser], [die] beiden [Brüder]
PPER	irreflexives Personalpronomen	ich, er, ihm, mich, dir
PPOSS	substituierendes Possessivpronomen	meins, deiner

<sup>1</sup> Das erweiterte Pos-Tagset des Gryphius-Projektes wird als Basis für das Dramen-Projekt übernommen und ggf. um weitere projektspezifische Tags ergänzt.

TP 4 | Personenreferierende Pronomen in Dramen

PPOSAT	attribuierendes Possessivpronomen	mein [Buch], deine [Mutter]
PRELS	substituierendes Relativpronomen	[der Hund,] der
PRELAT	attribuierendes Relativpronomen	[der Mann], dessen [Hund]
PRF	reflexives Personalpronomen	sich, einander, dich, mir
PWS	substituierendes Interrogativpronomen	wer, was
PWAT	attribuierendes Interrogativpronomen	Welche [Farbe], wessen [Hut]
PWAV	adverbiales Interrogativpronomen	warum, wo, wann, worüber, wobei
PAV	pronominaladverb	dafür, dabei, deswegen, trotzdem
PTKZU	„zu“ vor Infinitiv	zu [gehen]
PTKNEG	Negationspartikel	nicht
PTKVZ	Abgetrennte Verbzusatz	[er kommt] an, [er fährt] rad
PTKANT	Antwortpartikel	Ja, nein, danke, bitte
PTKA	Partikel bei Adjektiv oder Adverb	am [schönsten], zu [schnell]
TRUNC	Kompositions-Erstglied	An- [und Abreise]
VVFIN	Finites Verb, voll	[du] gehst [wir] kommen [an]
VVIMP	Imperativ, voll	Komm [!]
VVINFINF	Infinitiv, voll	gehen, ankommen
VVIZU	Infinitiv mit „zu“, voll	anzukommen, loszulassen
VVPP	Partizip Perfekt, voll	gegangen, angekommen
VAFIN	Finites Verb, aux	[du] bist, [wir] werden
VAIMP	Imperativ, aux	Sei [ruhig !]
VAINFINF	Infinitiv, aux	Werden, sein
VAPP	Partizip Perfekt, aux	gewesen
VMFIN	Finites Verb, modal	dürfen
VMINFINF	Infinitiv, modal	wollen
VMPP	Partizip Perfekt, modal	gekonnt
XY	Nichtwort, Sonderzeichen enthaltend	3:7, H2O, D2XW3
\$,	Komma	,
\$.	Satzbeendende Interpunktion	.?!;:
\$(	Sonstige Satzzeichen, satzintern	-[,]()